



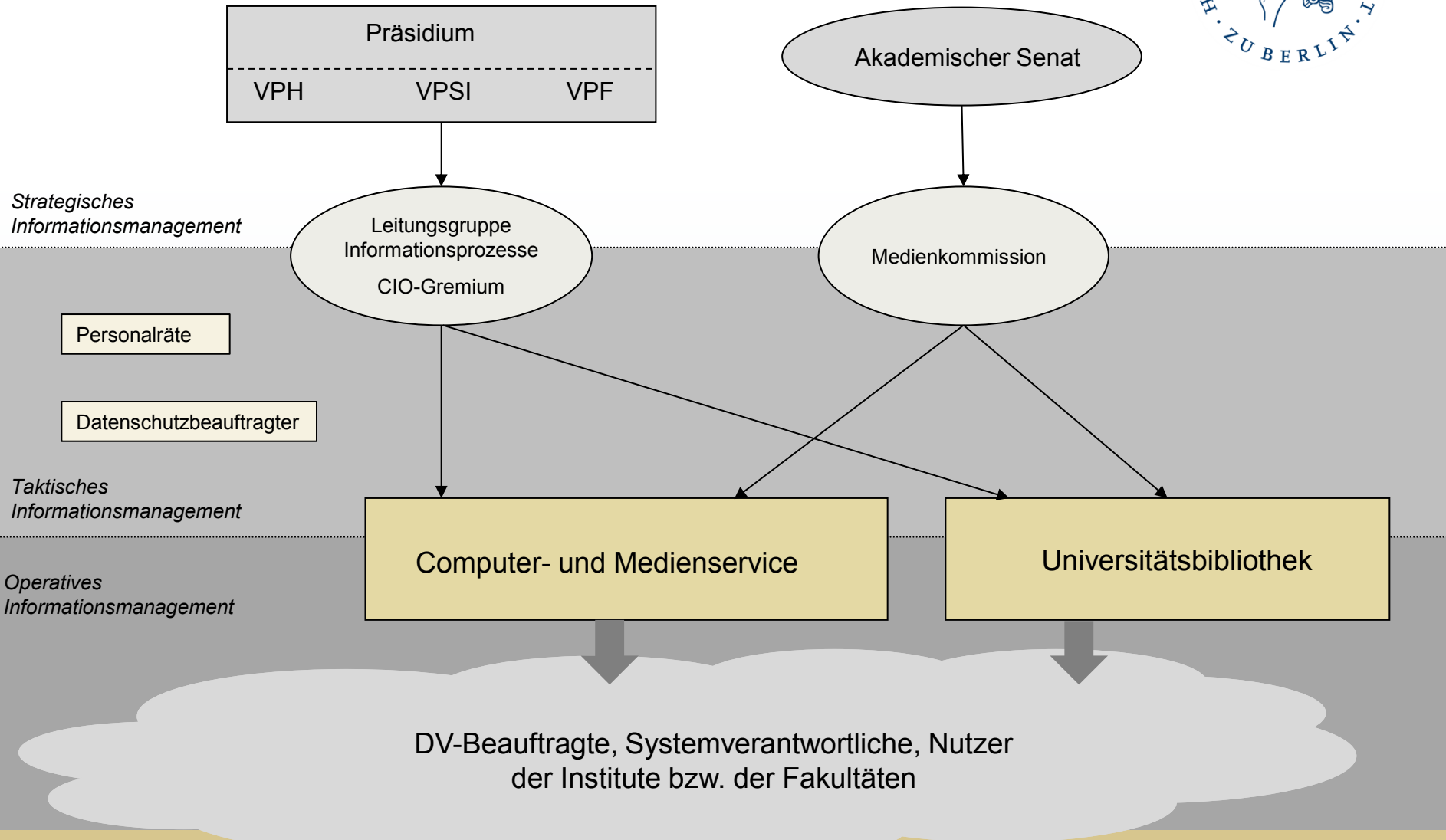
# Service für die Forschung an der Humboldt-Universität zu Berlin

Zentraleinrichtung  
Computer- und Medienservice  
Prof. Dr. Peter Schirmbacher  
[schirmbacher@cms.hu-berlin.de](mailto:schirmbacher@cms.hu-berlin.de)

## Agenda:

1. IT an der Humboldt-Universität
2. Infrastruktur
3. Service
4. Forschung
5. Ausblick

# IT-Organisation an der Humboldt-Universität



Vizepräsident für  
Forschung

Medienkommission  
des AS

Wissenschaftlicher Beirat

Wiss. Direktor / Techn. Direktor

Verwaltung / Sekretariat

Abt. System-  
software  
und  
Kommunikation



Abt. Hard-  
und  
Software-  
service



Abt. Multimedia-  
service



Abt. DV  
in der  
Verwaltung



CMS / UB AG  
Elektronisches  
Publizieren

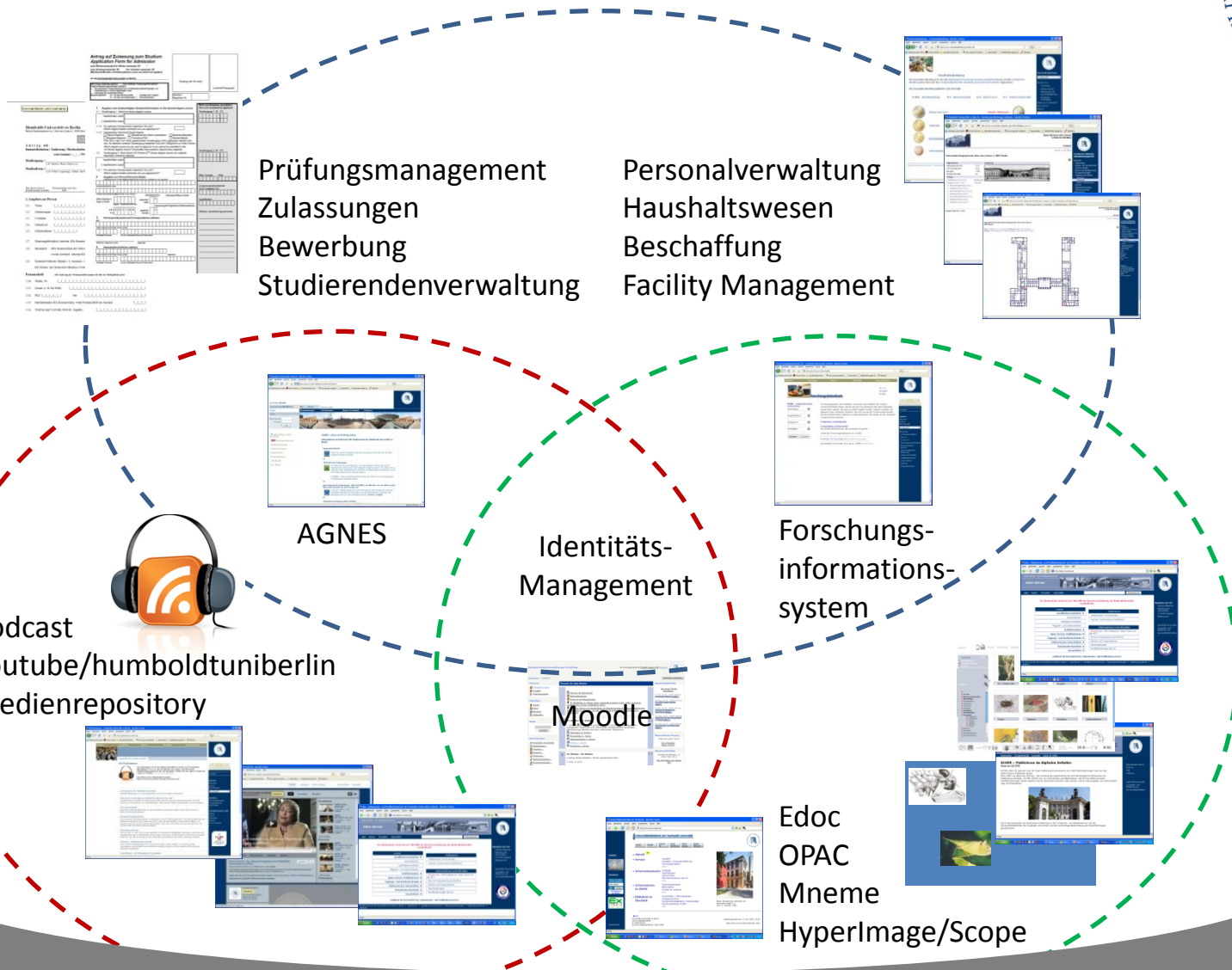


# Aufgaben des CMS

Die vorrangige Aufgabe des CMS ist die Unterstützung von Lehre, Studium, Forschung und Verwaltung durch IT-Dienstleistungen:

- Planung, Ausbau und Betrieb des Universitätsrechnernetzes und des Speichernetzes der HU sowie der Anschlüsse an äußere Netze
- Aufbau und Betrieb von zentralen, einrichtungsübergreifend nutzbaren Serverdiensten für Information, Kommunikation, Fileservice, Datenbanken, Datensicherung und Computeservice
- Projektierung und Betrieb zentralisierter Multimediaaustattung der HU
- Implementierung und Betrieb von IT-Anwendungen der Universitätsverwaltung
- Absicherung eines Hard- und Softwareservices sowie des Betriebs zentraler öffentlicher Computerarbeitsplätze für die HU
- Planung der Beschaffung sowie Auswahl zentral betriebener Informationstechnik, Beratung der Einrichtungen der HU bei der Planung und Auswahl von IT.

# Informationsinfrastruktur



**Basisdienste**

## Entwicklung der IT an der Humboldt-Universität

<b>Ausgewählte Kennzahlen</b>	<b>Ausstattung 1992</b>	<b>Ausstattung 2004</b>	<b>Ausstattung 2013</b>
Accountinhaber	~4.000	~38.000	Aktive Nutzer 34.000
Computerplätze	~2.000	~ 7.900	~ 11.000
Internetanschluss	9,6 Kbit/s	622 Mbit/s	4 Gbit/s
LAN Breitbandanschluss	~5	~30	~50
WLAN-AccessPoints		240	1.187
Archiv/Backup-Kapazität	500 Gbyte	150 Tbyte -netto	4.000 Tbyte -netto
Storage Area Network		30 Tbyte -netto	2.500 Tbyte -netto
Web-Server	3 Gopher-Server	255.000 Hits/Tag	2,27 Mio Hits/Tag
Moodle – Kurse		~ 3.800	22.300
aktive Nutzer		18.000	39.000
Verwaltungsanwendg.	~ 35	~ 96	~ 240
Rechner im Verwaltungsnetz	(Testnetz) 10	~ 440	~ 650

# 10 Thesen zur strategischen Planung des CMS

- Hochverfügbarkeit vs. infrastrukturelle und bauliche Voraussetzungen
- IT-Organisation und IT-Sicherheit an der HU
- Finanzielle und personelle Voraussetzungen
- Basisdienste und Innovation
- Entwicklung eines modernen Campusmanagementsystems
- E-Learning
- Unterstützung des Forschungsprozesses
- Elektronisches Publizieren
- Verwaltungsdatenverarbeitung
- Effizienz der eigenen Arbeit



# Infrastruktur



# Forschungsdaten

„Unter **digitalen Forschungsdaten** verstehen wir alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind. Der Forschungsprozess umfasst dabei den gesamten Kreislauf von der Forschungsdatengenerierung, z. B. durch ein Experiment in den Naturwissenschaften, eine dokumentierte Beobachtung in einer Kulturwissenschaft oder eine empirische Studie in den Sozialwissenschaften, über die Bearbeitung und Analyse bis hin zur Publikation und Archivierung von Forschungsdaten.“

„**Digitale Forschungsdaten entstehen in allen Wissenschaftsdisziplinen und unter Anwendung verschiedener Methoden, abhängig von der Forschungsfrage.** Dies hat zur Folge, dass sie in unterschiedlichen Medientypen, Aggregationsstufen und Datenformaten auftreten. Um das Bereitstellen von Forschungsdaten und ihre Nachnutzung zu ermöglichen, sind Metadaten und eine Datendokumentation essentiell, die den Kontext der Forschungsdaten beschreiben sowie die Werkzeuge, mit denen sie erzeugt, gespeichert, bearbeitet und analysiert wurden.“

[Quelle: Kindling, Maxi & Schirmbacher, Peter (2013) Die digitale Forschungswelt als Gegenstand der Forschung. In: Information, Wissenschaft & Praxis 2-3/2013, ]

Der Umgang mit Forschungsdaten hat drei Dimensionen:

## Wissenschaftspolitische Dimension

Wer ist verantwortlich und wie sehen die politischen, finanziellen und personellen Rahmenbedingungen regional, national und international aus.

## Organisatorische Dimension

Welche Strukturen werden geschaffen und wie kann eine entsprechende Arbeitsteilung zwischen den Einrichtungen und zwischen Wissenschaftlern und Servicepersonal aussehen.

## Technische Dimension

Auf welcher technischen Basis wird die Infrastruktur geschaffen. Wie wird ein entsprechender Service unter Einhaltung sämtlicher rechtlichen Rahmenbedingungen gewährleistet.

# Anforderungen an die Aufbewahrung wissenschaftlicher Daten



## 1. Zugänglichkeit

- für die Zielgruppe Nachnutzenden verfügbar und nutzbar
- für die Zielgruppe der Forschenden verfügbar und nutzbar

## 2. Nachhaltigkeit

- Dauerhafte Verfügbarkeit
- Nachweis der Unveränderbarkeit der Daten Zitierbarkeit

## 3. Nachvollziehbarkeit

- Nachvollziehbarkeit des gesamten Erhebungsprozesses
- Widerspiegelung des zeitlichen Ablaufs
- Dauerhafte Identifizierbarkeit

## 4. Authentizität

- Glaubwürdigkeit der Autorenschaft

# Anforderungen an die Aufbewahrung wissenschaftlicher Daten



## 5. Qualitätssicherung (vor oder mit der Ablieferung)

- Peer Review
- Plausibilitätskontrollen

## 6. Bewertung (nach der Ablieferung)

- Zitationsrate
- Nachnutzungsintensität

## 7. Geschwindigkeit

- Nachweisfunktion der Daten
- Aktualität der Daten

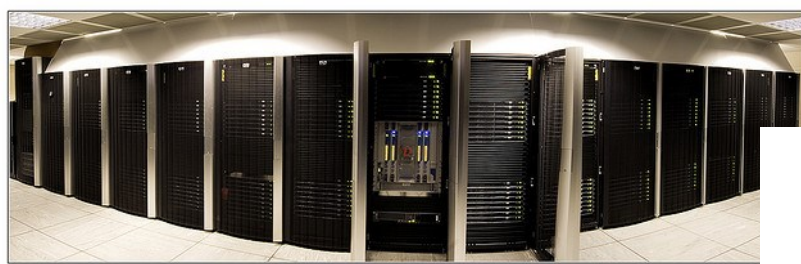
## 8. Vollständigkeit

- Beschreibung mit umfassenden Metadaten  
(Technische Metadaten, Deskriptive Metadaten, LZA Metadaten, ...)

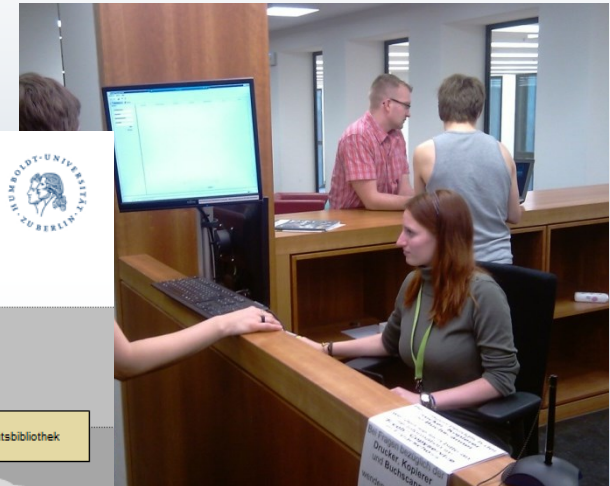
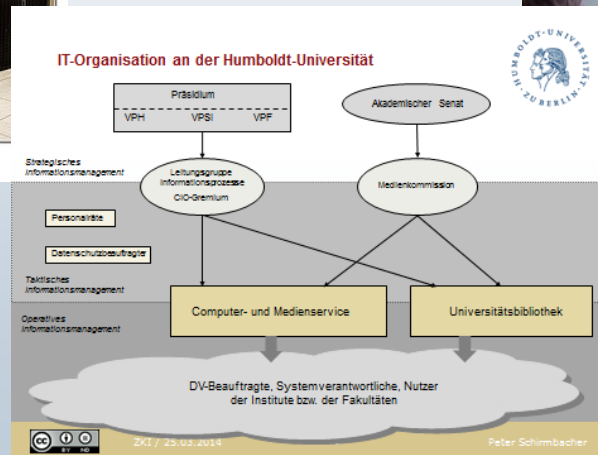


# Unter Informationsinfrastruktur

(für Wissenschaft und Forschung)  
 versteht man die Gesamtheit der  
*technischen,*  
*organisatorischen,*  
*finanziellen und*  
*personellen* Ressourcen  
 zur Gewährleistung der Informationsversorgung  
 des wissenschaftlichen Bereiches der Gesellschaft.



Jupiter E

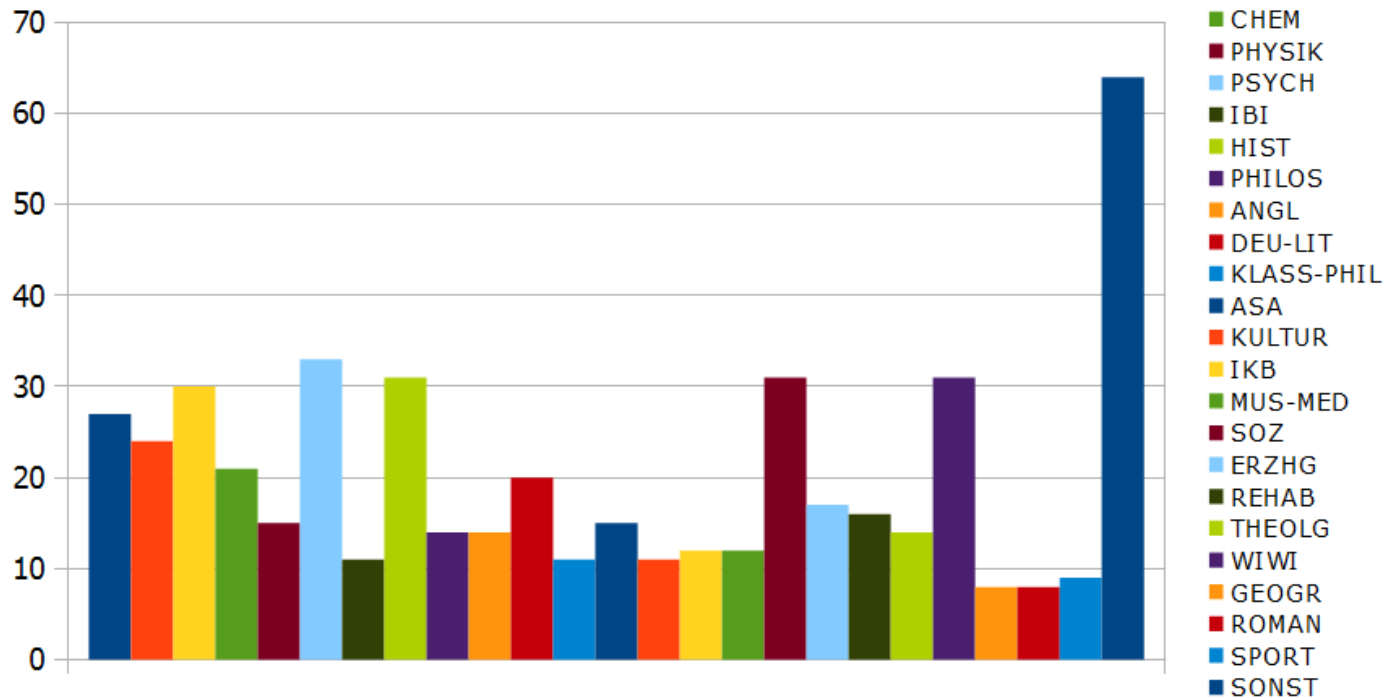


## Service



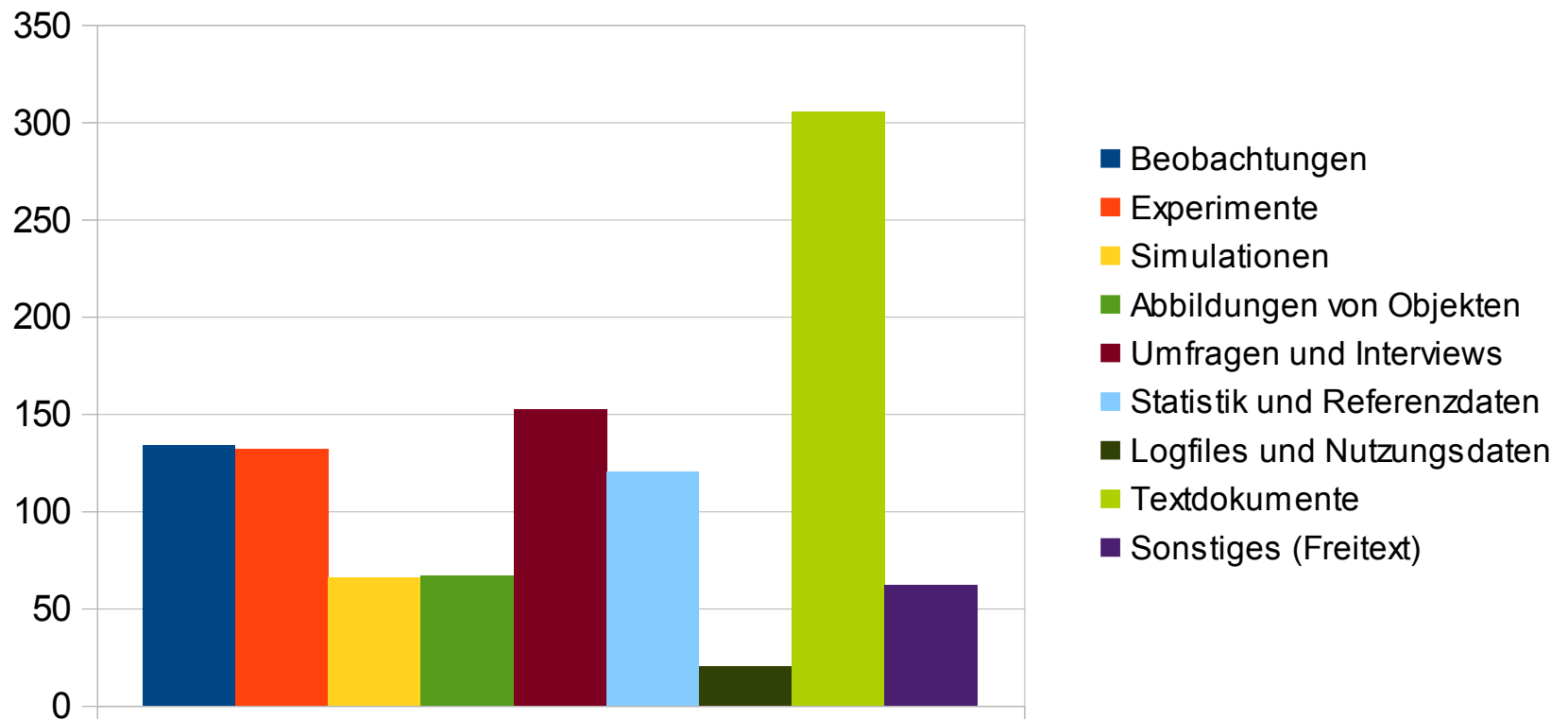
## Umfrage an der HU zum Umgang mit Forschungsdaten / 1. Quartal 2013

Verteilung der Teilnehmer nach Einrichtungen  
(n=499)

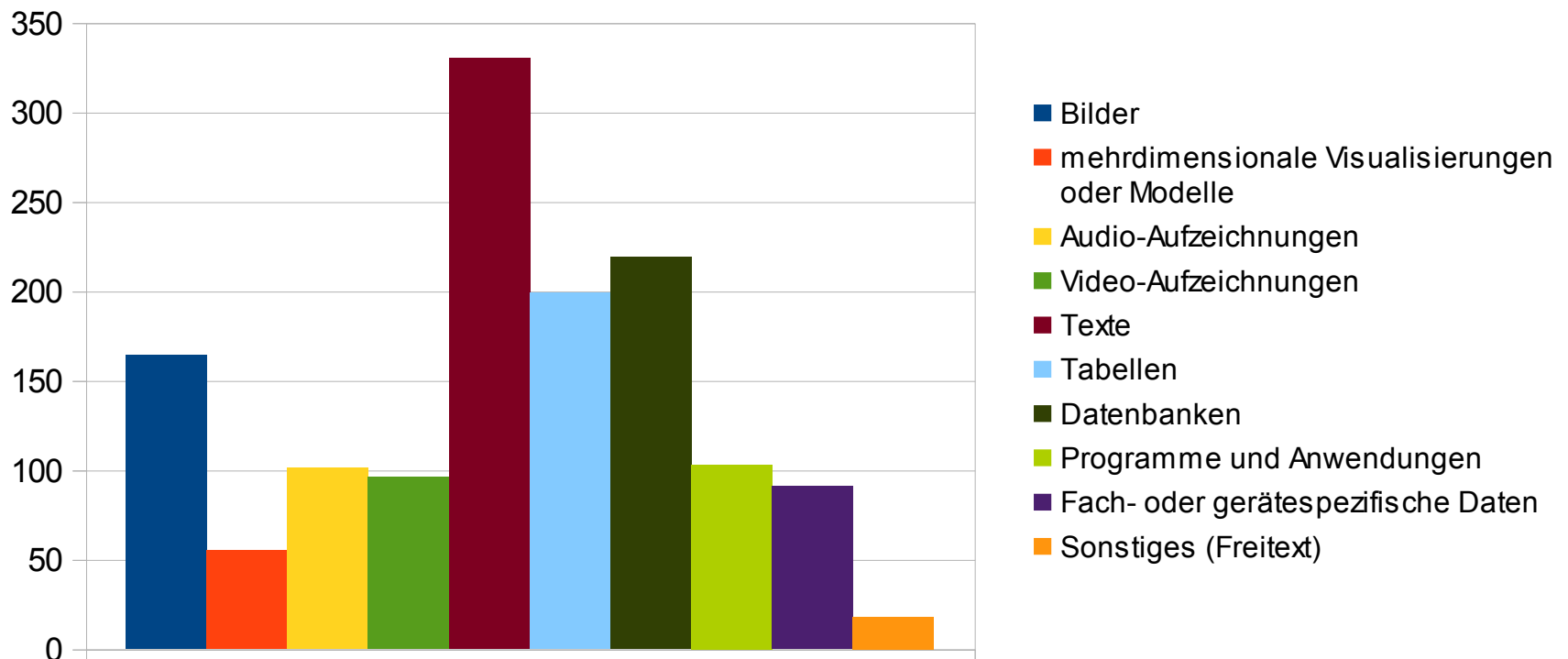




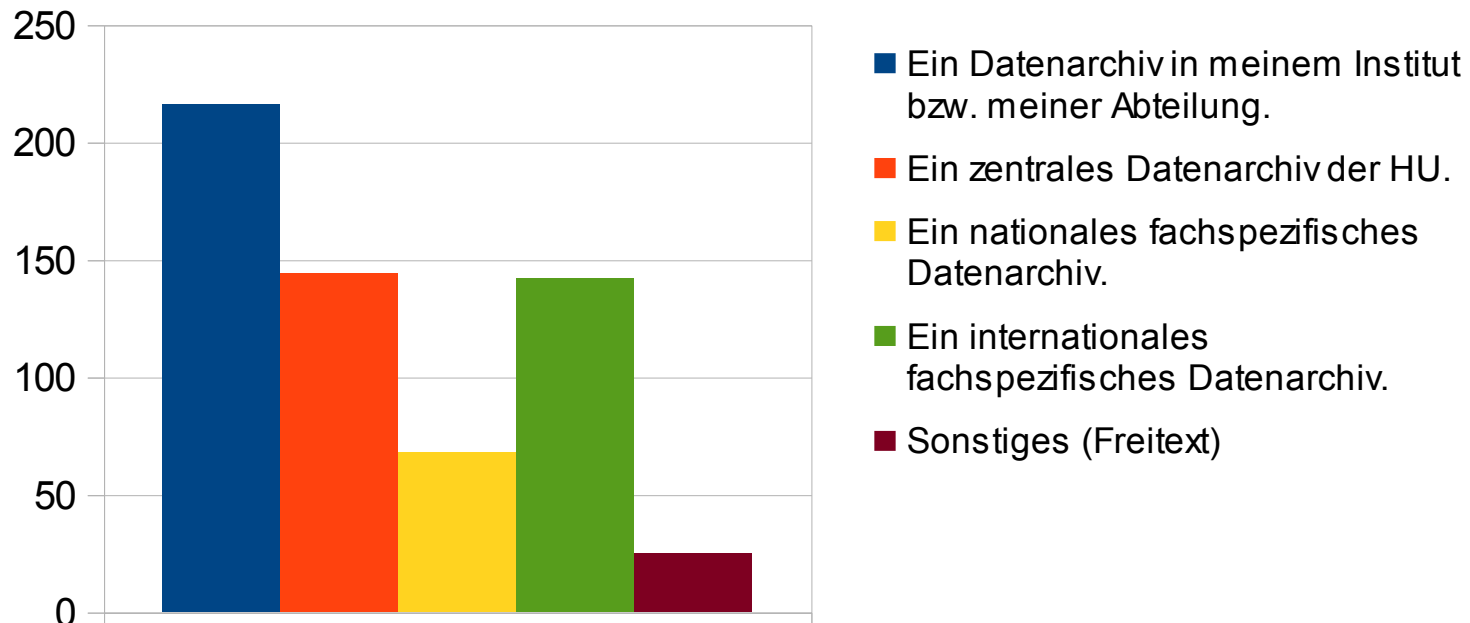
## Aus welchen Quellen schöpfen Sie Ihre Forschungsdaten hauptsächlich?



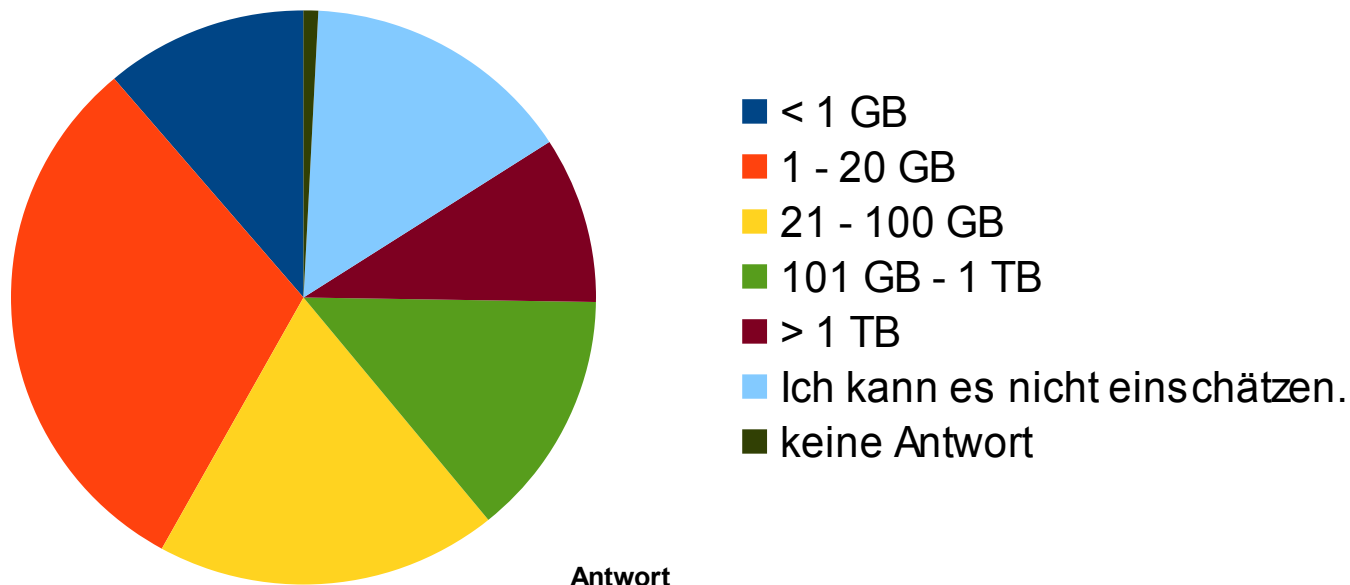
## Um welche Datentypen handelt es sich?



## Welches Datenarchiv würden Sie für die Ablage ihrer Daten am ehesten nutzen?

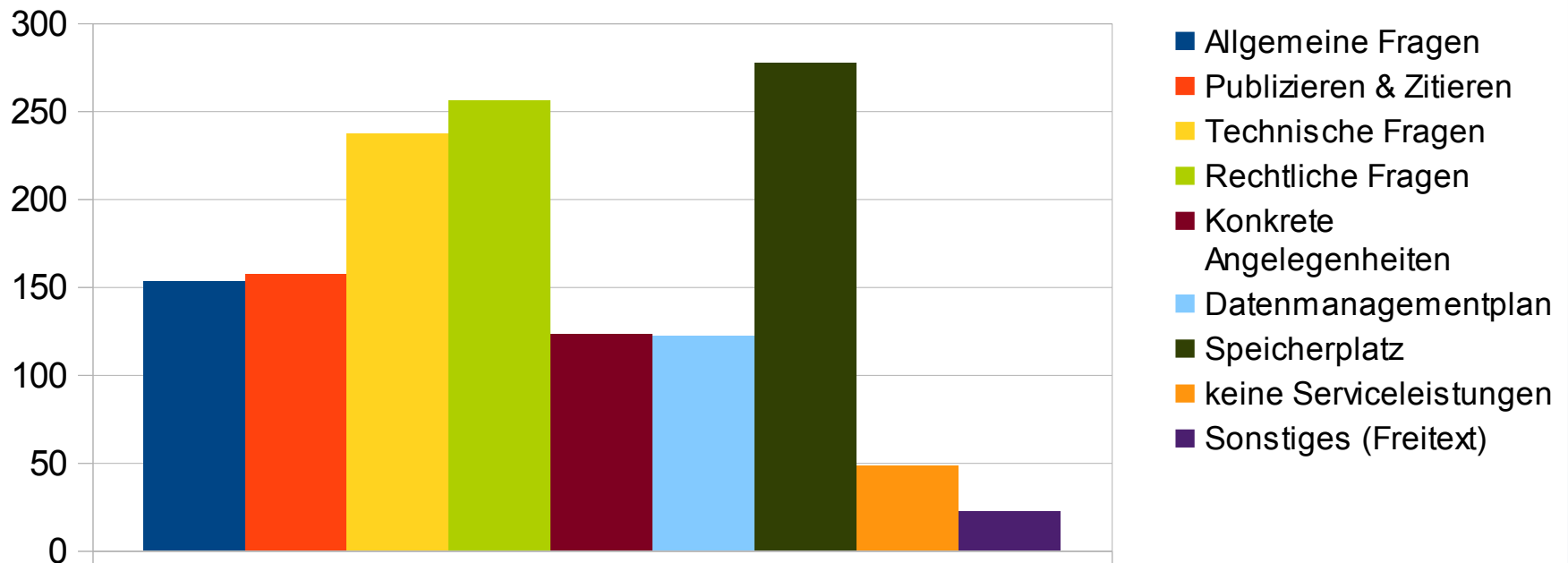


Bitte schätzen Sie die Gesamtgröße Ihrer Forschungsdaten bezogen auf den benötigten Speicherplatz.



Antwort	Anzahl	Prozent
< 1 GB	55	11.20%
1 - 20 GB	151	30.75%
21 - 100 GB	93	18.94%
101 GB - 1 TB	68	13.85%
> 1 TB	46	9.37%
Ich kann es nicht einschätzen.	74	15.07%
keine Antwort	4	0.81%

## Welche Serviceleistungen würden Sie sich von der HU wünschen? (Forschungsdatenmanagement)



## Policies zum Umgang mit Forschungsdaten

### Hauptaussagen einer Policies zum Management von Forschungsdaten:

(in enger Anlehnung an das DINI-Zertifikat für den Dokumenten- und Publikationsservice)

- Der Betreiber verfügt über öffentlich bereitgestellte Leitlinien (Policy), die den Dienst beschreiben.
- Eine Festlegung der Rechte und Pflichten des Betreibers des Repositoriums
- Eine Festlegung der Rechte und Pflichten der Produzenten / Ablieferer von Forschungsdaten.
- Eine Beschreibung der Art der Dokumente, die im Repository vorgehalten werden, sowie Anforderungen an deren inhaltliche und technische Qualität.
- Eine Festlegung darüber, wie lange die Forschungsdaten in dem Repository mindestens verfügbar gehalten werden, und die damit verbundene Garantieerklärung.
- Aussagen zum technischen Betrieb des Repositoriums.

# Grundsätze zum Umgang mit Forschungsdaten an der HU



1. Forschende HU-Angehörige sind verpflichtet, die Forschungsdaten sicher zu **speichern, angemessen aufzubereiten und zu dokumentieren sowie langfristig aufzubewahren**. Die Verantwortung für die Gewährleistung dieser Prozesse liegt bei den HU-Angehörigen, die das Forschungsvorhaben leiten.
2. Alle forschenden HU-Angehörigen sind aufgefordert, die in ihrer wissenschaftlichen Tätigkeit entstehenden Forschungsdaten gemäß den im jeweiligen Fachgebiet **etablierten Regelungen bzw. Standards** aufzubereiten. Sie dokumentieren den **gesamten Forschungszyklus** sowie die verwendeten Werkzeuge und Verfahren.

# Grundsätze zum Umgang mit Forschungsdaten an der HU



3. Es liegt in **eigener Verantwortung** der forschenden HU-Angehörigen, zu welchem Zeitpunkt und zu welchen rechtlichen Bedingungen Forschungsdaten **zugänglich gemacht werden**. Die Humboldt-Universität empfiehlt, Forschungsdaten ebenso wie die wissenschaftliche Publikation gemäß der **Open-Access-Erklärung** der HU öffentlich zugänglich zu machen. Der Schutz personenbezogener Daten, des Urheberrechts und der berechtigten Interessen Dritter muss gewährleistet sein.
4. Die Forschungsdaten, die Grundlage einer Publikation bilden, sollen **langfristig in einem geeigneten vertrauenswürdigen Datenarchiv bzw. Repository archiviert und/oder veröffentlicht** werden. Sie zählen zum wissenschaftlichen Output der Forschenden der Humboldt-Universität.



# Handlungsempfehlungen

*in Ergänzung zu den Grundsätzen zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin*

## I. Planungsphase: vor dem Forschungsvorhaben

- Anforderungen der Förderorganisationen  
z.B. Forschungsdatenmanagementplan
- Rechtliche Fragen

## II. Implementierungsphase: während des Forschungsvorhabens

- Einhaltung fachspezifischer Standards

## III. Abschlussphase: nach dem Forschungsvorhaben:

- Zugänglichmachung von Forschungsdaten

# Forschung



# Laudatio

Long Term Access and Usage of Deeply Annotated Information

**Eine** Möglichkeit Forschungsdaten (historische Textdaten) für eine Fachdisziplin (historische Linguistik) nachhaltig und nutzerorientiert zu speichern und OA (wieder)- nutzbar zu machen.

Erste Ausschreibung der DFG im Bereich **Infrastrukturen für Forschungsdaten.**



# Laudatio

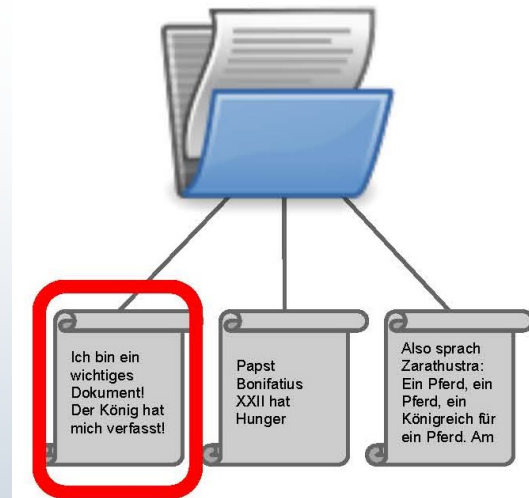
## Ziele und Herausforderungen

- Fokus: deutsche historische Texte und deren linguistische Annotationen
- Langfristige Speicherung und (Wieder)- Verwendung von historischen Korpora
- Zugang zu historischen Korpora
- Darstellung mittels tief strukturierter Metadaten
- Suche nach Korpora und deren Dokumenten
- Begriffsverständnis von Forschungsdaten
- Geringe Verbreitung von Standards der Datenaufbereitung und Speicherung
- Einbindung weiterer Fachdisziplinen wie z.B. Musik-, Geschichts- und Literaturwissenschaftler, die auf altdeutschen Texten arbeiten
- Nachnutzung und Verfügbarkeit einer disziplinübergreifenden FDR-Lösung als Serviceangebot

# Laudatio

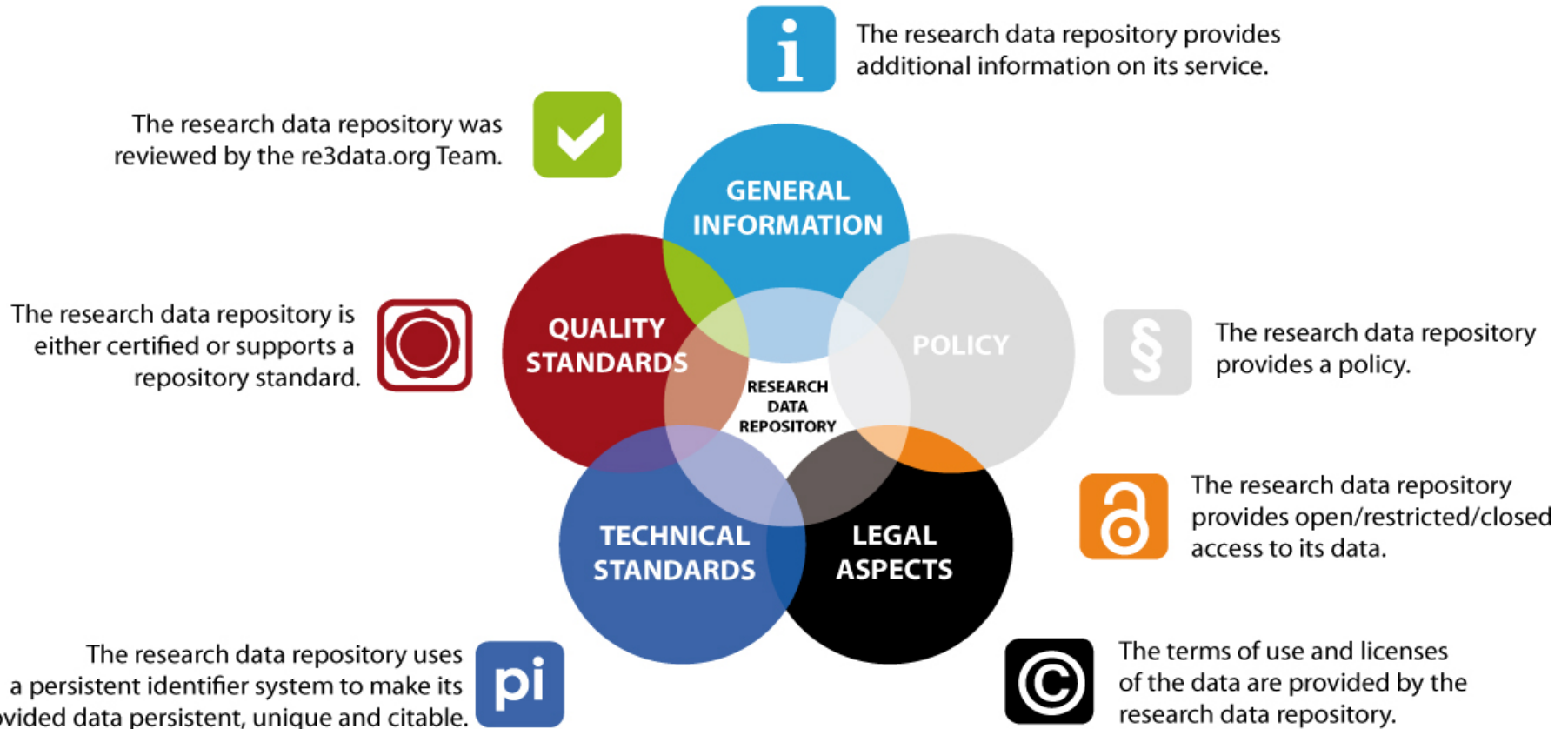
## Datenstruktur

- Ein Korpus besteht aus einem oder mehreren **Dokumenten**
- Ein Dokument besteht aus einer oder mehreren **Annotationsen**
- Es gibt einen oder mehrere **Autoren**, die das Dokument verfasst haben
- Es gibt einen oder mehrere **Editoren**, die das Korpus zusammenstellen
- Es gibt einen oder mehrere **Annotatoren**, die das Korpus bearbeiten
- **Annotationsen sind die primären Forschungsdaten**
- Beschreibung des Korpus mithilfe des TEI-Metadatenstandards





## 3. Organisatorische Dimension



## Search for repositories (alpha version)

mineralogy Search

Filter results »

Subject: Add Subjects ▼ Content Type: Add Types ▼ Country: Add Countries ▼

☐ Certificates
 ☐ Open Access
 ☐ Persistent Identifier
 ☒ Repository reviewed by re3data.org

remove filters

## Search Results (4 results)

Search terms: *mineralogy*

### Crystallography Open Database



Physics

Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers

### Inorganic Crystal Structure Database



Chemistry Geochemistry Materials Science and Engineering Mineralogy and Crystallography Physics

The most comprehensive database on fully determined inorganic crystal structures • Full structural data: cell parameters, atom positions for all entries, displacement parameters • Full bibliographic data: publication title, journal reference(s), author names • Full structure description: Structural formula, compositions, ANX formulae, structure types • High-quality data: extensive data evaluation and correction by senior experts • Web and PC based software solutions, data updated twice a year • 25+ years of serving the scientific community

### PANGAEA



Biology Geochemistry Geodesy Geoinformatics Geology and Palaeontology Geophysics Mineralogy and Crystallography Oceanography Remote Sensing

The Information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.

# re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

[Home](#) [Search](#) [Suggest](#) [FAQ](#) [About](#) [Schema](#) [Contact](#) [Imprint](#)

## Search for repositories (alpha version)

 [Filter results](#)

## Search Results (173 results)

### CLARIN



#### Linguistics

CLARIN aims at uniting existing digital archives in Europe that contain language-based material into a federation that will allow the social sciences and humanities research communities to have unified access to the content. It wants to make the wealth of language and speech processing tools that have been developed over the recent years available to interested researchers with a view to opening up new research avenues. Another goal of CLARIN is to provide web based services that will allow non-expert users (especially humanities and social sciences researchers without technological background) to perform complex tasks on the materials contained in the archives, such as 'Summarize Le Monde of March 17 2008 - in Polish'. Only countries and institutions can be CLARIN members. Private persons are not allowed.

### DANS



#### Ancient Cultures Geosciences (including Geography) History Social and Behavioural Sciences

DANS promotes sustained access to digital research data. For this purpose, DANS encourages researchers to archive and reuse data in a sustained manner, e.g. through the online archiving system EASY. DANS also provides access, via NARCIS.nl, to thousands of scientific datasets, e-publications and other research information in the Netherlands. In addition, the Institute provides training and advice, and performs research into sustained access to digital information. Driven by data, DANS ensures that access to digital research data keeps improving, through its services and by taking part in national & international projects and networks. DANS is an Institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO).

### Ecological Society of America esa Ecological Archives



#### Agriculture Biology Chemistry Forestry Horticulture and Veterinary Medicine Life Sciences





**Danke für die  
Aufmerksamkeit**